JOURNAL OF COMPUTER AND SYSTEM SCIENCES 26, 82–91 (1983)

# Checking Sets, Test Sets, Rich Languages and Commutatively Closed Languages*

JÜRGEN ALBERT

*Institut für Angewandte Informatik und Formale Beschreibungsverfahren, Universität Karlsruhe, D–7500 Karlsruhe, West Germany*

AND

DERICK WOOD

*Department of Computer Science, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada*

Received April 7, 1981; revised September 1, 1982

The problem of homomorphism equivalence is to decide for some language $L$ over some finite alphabet $\Sigma$ and two homomorphisms $f$ and $g$ whether or not $f(x) = g(x)$ for all $x$ in $L$. It has been conjectured that each $L$ can be represented by some finite subset $F$ such that for all pairs of homomorphisms $f$ and $g$: $f(x) = g(x)$ for all $x$ in $F$ implies $f(x) = g(x)$ for all $x$ in $L$. This conjecture is proved for the families of rich and commutatively closed languages. Lower and upper bounds are derived for the sizes of these finite subsets and examples of language families are given for which there are effective constructions of these subsets.

## 1. INTRODUCTION

Although a homomorphism can be considered to be the simplest function which translates a word (letter by letter) into some other word, many basic problems concerning pairs of homomorphisms are still open. For example, given two homomorphisms $f$ and $g$ over some alphabet $\Sigma$, does there exist some $x$ in $\Sigma^+$ such that $f(x) = g(x)$? This is a reformulation of the well-known Post correspondence problem, however, there are some recent results about the decidability of restricted versions of it. Despite this the minimal size of $\Sigma$ such that the Post correspondence problem over $\Sigma$ becomes undecidable remains open.

We can use two homomorphisms in this way to define a language, their so-called *equality set*. For two homomorphisms $f$ and $g$ their equality set $E(f, g)$ is defined as $\{x \text{ in } \Sigma^* \mid f(x) = g(x)\}$. The power of this mechanism has been demonstrated by yielding elegant characterizations for the recursively enumerable sets and very general families of complexity classes including NP [4].

* This work was partially supported by grants from the Deutsche Forschungsgemeinschaft (DFG) and from the Natural Sciences and Engineering Research Council of Canada Grant No A–7700.



0022-0000/83/010082–10$03.00/0

The study of test sets can be considered as dual to this approach. Whereas for equality sets, we fix a pair of homomorphisms and generate a language from them, we now fix a language and all those pairs of homomorphisms are considered which have the same images for all words in the language, i.e., all homomorphisms $f$, $g$ for which $E(f, g)$ is a superset of the given language. To test this property for some $f$ and $g$ it would suffice for our language $L$ to always be effectively represented by some finite subset $F \subseteq L$. Such an $F$ is then called a *test set* for $L$. Test sets can be constructed effectively for regular and context-free languages and are known to exist for all languages over a binary alphabet [1, 4]. As applications of these results, we obtain algorithms to decide the equivalence of deterministic gsm mappings on context-free languages and theorems about the reducibility of systems of string equations of certain types [1, 6 and 13].

In this article, we treat the problem of test set existence for the families of rich [7] and commutatively closed [10] languages. It turns out that in both cases the characterization of the languages by their sets of Parikh vectors is sufficient in a sense to be made precise later. Before proving our results we need some notation.

Let $L$ be a language over some alphabet $\Sigma$ and $g$ and $h$ be two homomorphisms defined on $\Sigma^*$. We say *$g$ and $h$ agree on $L$*, denoted by $g \equiv_L h$, if for all words $x$ in $L$, $g(x) = h(x)$. We say that a finite set $F \subseteq \Sigma^*$ is a *checking set for $L$* iff for all pairs of homomorphisms $g$ and $h$ defined on $\Sigma^*$, $g$ and $h$ agree on $F$ iff $g$ and $h$ agree on $L$. If further $F \subseteq L$, then we say that $F$ is a *test set for $L$*. These notions were first introduced by Culik, II and Salomaa [8] and have been subsequently studied in [1, 6, 7, and 9]. Reference [4] contains a survey of recent results.

Associated with the notion of a test set are three fundamental problems. These are:

(i) The Test Set Existence Problem. Does every language have a test set? Ehrenfeucht has conjectured that this is indeed the case.

(ii) The Test Set Construction Problem. Given an arbitrary language can its test set be constructed effectively, if it has one? In general this is surely not the case, however, [1] demonstrates effectiveness for context-free languages.

(iii) The Test Set Decision Problem. Given an arbitrary language $L \subseteq \Sigma^*$ and an arbitrary finite set $F \subseteq \Sigma^*$, then is $F$ a test set for $L$?

In this article, we consider these three questions for the families of rich languages (Section 2) and commutatively closed languages (Section 3). A language $L \subseteq \Sigma^*$ is *rich* iff for all homomorphisms $g$ and $h$ such that $g(x) = h(x)$ for all $x$ in $L$, then $g$ and $h$ are identical. We say that $L$ is *commutatively closed* if for every word $x$ in $L$, $L$ contains every word $y$ in $\Sigma^*$ which has the same Parikh vector.

We are able to solve the existence problem affirmatively for these languages and also give partial results for the two other problems. Our approach also leads to another interesting result, namely, a necessary condition for a set to be a checking set and hence a test set for an arbitrary language. This also gives a lower bound on the size of a test set. We are also able to show that every language which has a checking set has a checking set of size given by this lower bound result. Moreover, Ehrenfeucht

has conjectured that an upper bound on the size of test sets is $2^n$, where $n$ is the cardinality of the alphabet. In the case of commutatively closed languages, we are able to derive an upper bound of $2^n(n! + n) + 5n^2$.

## 2. Checking and Test Sets for Languages and Rich Languages

The notions and definitions introduced here are mostly standard and can be found in many textbooks on formal language theory, e.g., [11, 12, and 14]. Let $\#_a(w)$ denote the number of occurrences of the letter $a$ in a word $w$ and alph($w$) the set of letters occurring in $w$. If $\Sigma = \{a_1,..., a_n\}$ is an alphabet and $w$ is in $\Sigma^*$, the $n$-dimensional vector $\mathbf{p}(w) = (\#_{a_1}(w), \#_{a_2}(w),..., \#_{a_n}(w))$ is called the Parikh vector of $w$.

In the following we shall use in the proofs some basic facts about combinatorial properties of words. Readers unfamiliar with these results are referred to [11, Sect. 11.3] for background material. We begin with a general lemma giving a necessary condition for a finite set to be a checking set for an arbitrary language.

Lemma 1. *Let $\Sigma = \{a_1,..., a_n\}$ be an alphabet and $L \subseteq \Sigma^*$. If $F$ is a checking set for $L$, then there are $y_1,..., y_m$ in $F$, $m \leqslant n$, such that for each $x$ in $L$ there exist rational numbers $\alpha_1,..., \alpha_m$ with $\mathbf{p}(x) = \sum_{i=1}^{m} \alpha_i \cdot \mathbf{p}(y_i)$.*

*Proof.* Let $y_1,..., y_m$ be in $F$ such that $\mathbf{p}(y_1),..., \mathbf{p}(y_m)$ constitutes a basis for $\mathbf{p}(F)$. We now show that there is a pair of homomorphisms $h_1$, $h_2: \Sigma^* \to \{a\}^*$ such that $h_1 \equiv_F h_2$ and $h_1(x) \neq h_2(x)$ if $\mathbf{p}(y_1),..., \mathbf{p}(y_m)$, $\mathbf{p}(x)$ are linearly independent.

Let $\mathbf{p}(x) = (x_1,..., x_n)$ and for $i = 1,..., m$ let $\mathbf{p}(y_i) = (y_{i1},..., y_{in})$. If $\mathbf{p}(y_1),..., \mathbf{p}(x)$ are linearly independent, the system of equations

$$\begin{aligned} &y_{11} \times \Delta_1 + y_{12} \times \Delta_2 + \cdots + y_{1n} \times \Delta_n = 0, \\ &\vdots \\ &y_{m1} \times \Delta_1 + y_{m2} \times \Delta_2 + \cdots + y_{mn} \times \Delta_n = 0, \\ &x_1 \times \Delta_1 + x_2 \times \Delta_2 + \cdots + x_n \times \Delta_n = 1, \end{aligned}$$

has at least one solution $(\Delta_1,..., \Delta_n)$ of rational numbers $\Delta_i$ by the basic theorems on the rank of such systems.

Let $\Delta_i = \alpha_i/\beta_i$, where $\alpha_i$ is in $\mathbb{Z}$, and $\beta_i$ is in $\mathbb{N}$, for $i = 1,..., n$. Now we complete the proof by exhibiting homomorphisms $h_1$ and $h_2$ such that $h_1 \equiv_F h_2$ and $|h_1(x)| - |h_2(x)| \neq 0$. To this end let $k = \beta_1 \times \beta_2 \times \cdots \times \beta_n$, choose natural numbers $\sigma_i$, $\tau_i$ such that $\sigma_i - \tau_i = k \times \Delta_i$ and define $h_1(a_i) = a^{\sigma_i}$, $h_2(a_i) = a^{\tau_i}$ for $i = 1,..., n$. Clearly, by the system of equations,

$$|h_1(y_i)| - |h_2(y_i)| = y_{i1} \times (\sigma_1 - \tau_1) + \cdots + y_{in} \times (\sigma_n - \tau_n) = 0$$

for $i = 1,..., m$ and

$$|h_1(x)| - |h_2(x)| = k \neq 0. \quad \blacksquare$$

An immediate corollary to this lemma gives a lower bound on the size of checking sets.

COROLLARY 2. *Let $L$ be an arbitrary language with a checking set $F$ and let a basis for $\mathbf{p}(L)$ have size $m$. Then $\#F \geqslant m$.*

We now state a trivial result which has far reaching consequences.

LEMMA 3. *Let $L \subseteq \Sigma^*$ be an arbitrary language and $F \subseteq \Sigma^*$ be an arbitrary finite language. Then $F$ is a checking set for $L$ iff $F$ is a checking set for $L^*$. Moreover if $F$ is a test set for $L$, then it is a test set for $L^*$.*

*Proof.* Trivial, since (i) for homomorphisms $h_1$, $h_2$ and words $x_1,..., x_k$ in $L$, $h_1(x_i) = h_2(x_i)$ implies $h_1(x_1 \ldots x_k) = h_2(x_1 \ldots x_k)$ and (ii) $L \subseteq L^*$. $\blacksquare$

DEFINITION. Let $L \subseteq \Sigma^*$ be an arbitrary language and $F \subseteq \Sigma^*$ be a checking set for $L$. We say $F$ is *minimal* if $\#F$ is the size of a basis for $\mathbf{p}(L)$. We define *minimal test sets* analogously.

THEOREM 4. *Let $L \subseteq \Sigma^*$ be an arbitrary language having a checking set. Then $L$ has a minimal checking set. If $L$ has a test set, then $L^*$ has a minimal test set.*

*Proof.* We shall only prove the first statement. The second follows from it together with Lemma 3.

Let $F \subseteq \Sigma^*$ be a checking set for $L$ and let $F = \{x_1,..., x_m, w_1,..., w_n\}$, where $\{\mathbf{p}(x_i): 1 \leqslant i \leqslant m\}$ forms a basis for $\mathbf{p}(L)$. If $n = 0$, then $F$ is already minimal, therefore consider $n > 0$. Now $\mathbf{p}(w_n) = \sum_{i=1}^{m} \alpha_i \times \mathbf{p}(x_i)$ for rational $\alpha_i$, not all which are negative. Hence without loss of generality assume $\alpha_1 \geqslant 0$.

Let $F' = \{x_1 w_n, x_2,..., x_m, w_1,..., w_{n-1}\}$. We shall show that $F'$ is a checking set for $F$ and hence for $L$. It is well known that $\{\mathbf{p}(x_1 w_n), \mathbf{p}(x_2),..., \mathbf{p}(x_m)\}$ is also a basis for $\mathbf{p}(L)$. Consider $h_1$ and $h_2$ with $h_1 \equiv_{F'} h_2$. We shall show that $h_1 \equiv_F h_2$. In other words, we shall show that $h_1(x_1 w_n) = h_2(x_1 w_n)$ implies $h_1(x_1) = h_2(x_1)$ and hence $h_1(w_n) = h_2(w_n)$. Now

$$\begin{aligned}
&\mathbf{p}(h_1(x_1 w_n)) - \mathbf{p}(h_2(x_1 w_n)) \\
&\quad = \mathbf{p}(h_1(x_1)) + \mathbf{p}(h_1(w_n)) - \mathbf{p}(h_2(x_1)) - \mathbf{p}(h_2(w_n)) \\
&\quad = (1 + \alpha_1)(\mathbf{p}(h_1(x_1)) - \mathbf{p}(h_2(x_1))) = 0
\end{aligned}$$

by the representation of $\mathbf{p}(w_n)$ given and because $x_1 w_n$, $x_2$, $x_3,..., x_m$ in $F'$ implies $\mathbf{p}(h_1(x_1 w_n)) = \mathbf{p}(h_2(x_1 w_n))$ as well as $\mathbf{p}(h_1(x_i)) = \mathbf{p}(h_2(x_i))$ for $i = 2, 3,..., m$. Now because $\alpha_1 \geqslant 0$, we have $\mathbf{p}(h_1(x_1)) = \mathbf{p}(h_2(x_1))$ and hence $|h_1(x_1)| = |h_2(x_1)|$. But this

means that in the equation $h_1(x_1 w_n) = h_2(x_1 w_n)$ which can be written as $h_1(x_1)\, h_1(w_n) = h_2(x_1)\, h_2(w_n)$ we have $h_1(x_1) = h_2(x_1)$. Therefore $h_1(w_n) = h_2(w_n)$ also. We have replaced $F$ by a checking set $F'$ satisfying $\#F' < \#F$. Clearly, this procedure can be iterated to obtain a minimal checking set. ∎

*Remark.* In [7], it is shown that the language $L = \{a^n b^n \mid n > 1\}$ cannot have a test set consisting of only one word, i.e., there exist languages which do not possess minimal test sets.

As an application of these results, we consider the collection of rich languages. A language $L \subseteq \Sigma^*$ is *rich* if for all homomorphisms $g$ and $h$ satisfying $g \equiv_L h$, we have $g \equiv_{\Sigma^*} h$.

THEOREM 5. *Let $\Sigma = \{a_1, \ldots, a_m\}$ and $L \subseteq \Sigma^*$. Then $L$ is rich iff there exist $x_1, \ldots, x_m$ in $L$ such that $\mathbf{p}(x_1), \ldots, \mathbf{p}(x_m)$ are linearly independent.*

*Proof.* ($\rightarrow$). Since $L$ is rich it must contain, by the arguments in Lemma 1, words $x_1, \ldots, x_m$ with $\mathbf{p}(x) = \sum_{i=1}^{m} \alpha_i \times \mathbf{p}(x_i)$, for some $\alpha_i$, for all $x$ in $\Sigma^*$. (Essentially $L$ is a, possibly infinite, checking set for $\Sigma^*$.)

($\leftarrow$). Each $\mathbf{p}(a_i)$ can be expressed as $\alpha_{i1} \times \mathbf{p}(x_1) + \cdots + \alpha_{im} \times \mathbf{p}(x_m)$. Whenever $h_1 \equiv_L h_2$ for two homomorphisms $h_1$, $h_2$, it follows that $|h_1(a_i)| = |h_2(a_i)|$ for all $i$ in $\{1, \ldots, m\}$. This immediately implies $h_1(a_i) = h_2(a_i)$ for all $i$ and thus $h_1 \equiv_{\Sigma^*} h_2$.

COROLLARY 6 (Test set existence). *Every rich language has a test set and moreover it has a minimal test set of the same size as its alphabet.*

COROLLARY 7 (Test set construction). *Given an arbitrary rich context-free language $L$ a test set for $L$ can be found effectively.*

This follows from the fact that $\mathbf{p}(L)$ is semilinear when $L$ is context free and hence a basis for $\mathbf{p}(L)$ can be found effectively. This result can be strengthened by observing that richness is decidable for context-free languages.

COROLLARY 8 (Test and checking set decision). *Given an arbitrary rich context-sensitive language $L \subseteq \Sigma^*$ and an arbitrary finite set $F \subseteq \Sigma^*$ it is decidable whether or not $F$ is a test set for $L$. Given an arbitrary rich language $L \subseteq \Sigma^*$ and an arbitrary finite set $F \subseteq \Sigma^*$ it is decidable whether or not $F$ is a checking set for $L$.*

Corollary 8 follows by observing that a finite set $F$ is a checking set for an arbitrary rich language iff $F$ is rich itself. Thus, testing whether $F$ is a test set for $L$ involves checking whether or not $\mathbf{p}(F)$ contains $\#\Sigma$ linearly independent vectors and testing if $F \subseteq L$. The latter test is effective for context-sensitive languages.

*Remark.* Richness is undecidable for context-sensitive languages $L \subseteq \Sigma^*$, since there is no algorithm to find the minimal alphabet $\Sigma'$ such that $L \subseteq (\Sigma')^*$.

## 3. Commutatively Closed Languages

We define the *commutative closure of a language* $L \subseteq \Sigma^*$, denoted by $c(L)$, by: $\{x$ in $\Sigma^*$: $\mathbf{p}(x) = \mathbf{p}(y)$ for some $y$ in $L\}$. We say $L$ is *commutatively closed* if $L = c(L)$.

Since a commutatively closed language $L$ is in some sense representable by its set of Parikh vectors $\mathbf{p}(L)$ one is led to think that any basis of $\mathbf{p}(L)$ can be chosen as a test set for $L$. The following example demonstrates that this is, in general, not the case:

Example. Let $L = \{x$ in $\{a, b\}^*$: $\#_a(x) = \#_b(x)\}$. Then $F = \{ab, ba\}$ is *not* a test set for $L$. Consider $h_1, h_2$: $\{a, b\}^* \to \{0, 1)^*$ defined by:

$$h_1(a) = 010, \qquad h_2(a) = 0, \qquad h_1(b) = 1, \qquad h_2(b) = 101.$$

Then

$$h_1(ab) = 0101 = h_2(ab), \qquad h_1(ba) = 1010 = h_2(ba),$$

but

$$h_1(aabb) = 01001011, \qquad h_2(aabb) = 00101101.$$

The proof of our main theorem shows that $F = \{aabb, abab, abba, baab, baba, bbaa\}$ can be chosen as a test set for $L$, for example.

Definition. Let $L \subseteq \Sigma^*$ be any commutatively closed language and let $F \subseteq L$ be finite and commutatively closed. We say that $F$ has property (c1) if:

> For each $z$ in $L$ there exist $x_1, \ldots, x_m$ in $F$ and rational numbers $\alpha_1, \ldots, \alpha_m$ such that $\text{alph}(x_i) = \text{alph}(z)$ for $i = 1, \ldots, m$ and $p(z) = \sum_{i=1}^{m} \alpha_i \times p(x_i)$,

and $F$ has property (c2) if:

> For each $z$ in $L$ there exist $x_1, \ldots, x_m$ in $F$ and rational numbers $\alpha_1, \ldots, \alpha_m$ such that $\text{alph}(x_i) = \text{alph}(z)$ for $i = 1, \ldots, m$ and $\mathbf{p}(z) = \sum_{i=1}^{m} \alpha_i \times \mathbf{p}(x_i)$,

Theorem 9. *Let $L \subseteq \Sigma^*$ be a commutatively closed language and $F$ a finite commutatively closed subset of $L$ with properties* (c1) *and* (c2). *Then $F$ is a test set for $L$.*

*Proof.* Let $z$ be in $L - F$ and $h_1$, $h_2$ be two homomorphisms with $h_1 \equiv_F h_2$. For $\Delta = \text{alph}(z)$, we can assume that there is an $\alpha$ in $\Delta$ with $h_1(\alpha) \neq h_2(\alpha)$. Otherwise, $h_1(z) = h_2(z)$ holds trivially. The set $\Delta$ is now partitioned as follows:

Let $\Delta_1$ be the set of all letters in $\Delta$ which occur exactly once in all $y$ in $F$ satisfying $\text{alph}(y) = \Delta$. Let $\Delta_2 = \Delta - \Delta_1$.

For our proof that $h_1(z) = h_2(z)$, we show that—except for one trivial subcase—all homomorphic images of all letters of $\Delta$ commute.

Let us consider the following cases:

*Case* 1. There is an $a$ in $\Delta_2$ with $h_1(a) \neq h_2(a)$.

*Case* 2. There is a $b$ in $\Delta_1$ with $h_1(b) \neq h_2(b)$ and $h_1(c) = h_2(c)$ for all $c$ in $\Delta_2$.

*Case* 1. Let $w = h_1(a)$, $w' = h_2(a)$. By property (c2) of $F$ and because $F$ is commutatively closed there are words $aya$, $aay$ in $F$, where $\text{alph}(ay) = \Delta$ and

$$wh_1(y)\, w = w' h_2(y)\, w', \tag{1}$$

$$wwh_1(y) = w'w' h_2(y). \tag{2}$$

Without loss of generality we can assume $|w| > |w'|$. Since $w$, $w'$ are prefixes (suffixes) of the same word, we have $w = w'x = \bar{x}w'$ for some $x$ and $\bar{x}$ with $|x| = |\bar{x}| > 0$. By substituting for $w$ with $w'x$ and $\bar{x}w'$ in Eqs. (1) and (2) we obtain

$$xh_1(y)\, \bar{x} = h_2(y) \tag{1'}$$

giving

$$xw'xh_1(y) = w'xh_1(y)\, \bar{x}. \tag{2'}$$

By (2'), $xw'$ and $w'x$ are prefixes of the same word. Since they are of the same length, they must be equal, that is, $xw' = w'x$. Because $w = w'x = \bar{x}w'$, we also conclude $\bar{x} = x$. This gives two simpler equations,

$$xh_1(y)\, x = h_2(y), \tag{1''}$$

$$xh_1(y) = h_1(y)\, x. \tag{2''}$$

By a basic theorem on commuting words Eq. (2'') implies the existence of a nonempty word $u$ and numbers $i \geqslant 1, j \geqslant 0$ such that $x = u^i$ and $h_1(y) = u^j$ (cf. [11, pp. 9 and 12]). Choosing $u$ to be of minimal length, determines $u$ uniquely as the "primitive root" of $x$. Now if $y'$ is a word having the same Parikh vector as $y$, then $ay'a$, $aay'$ are in $F$ as well. As shown we derive $xh_1(y') = h_1(y')\, x$. Since $u$ has been chosen uniquely and $|h_1(y')| = |h_1(y)|$ it follows that $h_1(y') = h_1(y)$. Thus $h_1(y)$ is not changed if the letters in $y$ are permuted.

Now by (1'') $h_2(y) = xh_1(y)\, x = xh_1(y')\, x = h_2(y') = u^{2i+j}$. Let $b$ be in the $\text{alph}(y)$ and $h_1(b)$ be nonempty, then $h_1(b)\, h_1(\bar{y}) = h_1(\bar{y})\, h_1(b)$, where $aab\bar{y}$ and $aa\bar{y}b$ are in $F$ and $\text{alph}(ab\bar{y}) = \Delta$. Again by [11, pp. 9 and 12] there is a unique nonempty word $v$, the primitive root of $h_1(b)$, and numbers $i' \geqslant 1$, $j' \geqslant 0$ such that $h_1(b) = v^{i'}$ and $h_1(\bar{y}) = v^{j'}$. Since $h_1(b)\, h_1(\bar{y}) = v^{i'+j'} = u^j$ and $v$ and $u$ are primitive it follows that $v = u$ and $j = i' + j'$. Similarly if $h_2(c)$ is nonempty and $c$ is in $\Delta$, there is a number $k \geqslant 1$ such that $h_2(c) = u^k$.

Now define $r(d)$, $s(d) \geqslant 0$ such that $h_1(d) = u^{r(d)}$ and $h_2(d) = u^{s(d)}$ for each $d$ in $\Delta$ and for each $d$ in $\Sigma - \Delta$ define $r(d) = s(d) = 0$. Let $\bar{\mathbf{r}}$ and $\bar{\mathbf{s}}$ denote the row vectors

$$\bar{\mathbf{r}} = (r(a_1), r(a_2), \ldots, r(a_n)),$$

$$\bar{\mathbf{s}} = (s(a_1), s(a_2), \ldots, s(a_n)),$$

where $\Sigma = \{a_1, a_2, ..., a_n\}$. By property (c1) of $F$ there exist $y_1, y_2, ..., y_m$ in $F$ with $\text{alph}(y_i) = \Delta$ and rational numbers $\alpha_1, ..., \alpha_m$ such that $\mathbf{p}(z) = \sum_{i=1}^{m} \alpha_i \times \mathbf{p}(y_i)$. Thus $h_1(z) = u^{\sigma_1}$, $h_2(z) = u^{\sigma_2}$, where

$$\sigma_1 = \mathbf{p}(z) \times \bar{\mathbf{r}}^{\mathrm{T}} = \sum_{i=1}^{m} \alpha_i \times \mathbf{p}(y_i) \times \bar{\mathbf{r}}^{\mathrm{T}},$$

and

$$\sigma_2 = \mathbf{p}(z) \times \bar{\mathbf{s}}^{\mathrm{T}} = \sum_{i=1}^{m} \alpha_i \times \mathbf{p}(y_i) \times \bar{\mathbf{s}}^{\mathrm{T}},$$

where we use T to denote transpose. Because $y_i$ is in $F$ and $h_1(y_i) = h_2(y_i)$ it follows that $\mathbf{p}(y_i) \times \bar{\mathbf{r}}^{\mathrm{T}} = \mathbf{p}(y_i) \times \bar{\mathbf{s}}^{\mathrm{T}}$ for $i = 1, ..., m$ and thus $h_1(z) = h_2(z)$.

*Case* 2. There is a $b$ in $\Delta_1$ with $h_1(b) \neq h_2(b)$ and for all $c$ in $\Delta_2$ $h_1(c) = h_2(c)$. Let us first consider the following simple subcase:

*Subcase* 2.1 ($h_1(c) = h_2(c) = \lambda$ for all $c$ in $\Delta_2$). Because of (c2) there is a word $y$ in $F$ such that $h_1(z) = h_1(y) = h_2(y) = h_2(z)$.

*Subcase* 2.2. We have $h_1(b) \neq h_2(b)$ for some $b$ in $\Delta_1$, $h_1(c) = h_2(c)$ for all $c$ in $\Delta_2$ and there is an $a$ in $\Delta_2$ such that $h_1(a) \neq \lambda$. Without loss of generality assume that $|h_1(b)| > |h_2(b)|$ and that there are words *aaby*, *aayb*, *ayab* in $F$ with $\text{alph}(aby) = \Delta$ such that $h_1$ and $h_2$ agree on these words. For $w = h_1(a) = h_2(a)$ and $v = h_1(b)$, $v' = h_2(b)$, we have $v = v'x = \bar{x}v'$ for some $x, \bar{x} \neq \lambda$. As in Case 1 we can now derive from $h_1(t) = h_2(t)$ for $t = aaby$, *aayb*, *abay*, *ayab* that $xw = wx$, $\bar{x}w = w\bar{x}$ and therefore $x = \bar{x}$, and $xh_1(y) = h_1(y)x$ for all $y$ such that *aaby* is in $F$ and $\text{alph}(aby) = \Delta$. This implies again that all homomorphic images of all letters in $\Delta$ commute and $h_1(z) = h_2(z)$. ∎

It should be obvious that every commutatively closed language $L$ has a finite commutatively closed subset $F$ satisfying properties (c1) and (c2). Thus, Theorem 9 implies

COROLLARY 10 (Test set existence). *Every commutativly closed language has a test set.*

As an application of Theorem 9, we obtain explicit test sets for some special commutatively closed languages.

COROLLARY 11. *Let* $\Sigma = \{a_1, ..., a_m\}$ *be an alphabet and* $L = \{x \text{ in } \Sigma^* : \#_{a_1}(x) = \#_{a_2}(x) = \cdots = \#_{a_m}(x)\}$. *Then for each* $i \geqslant 2$, $F_i = \{x \text{ in } L : \#_{a_j}(x) = i \text{ for } j = 1, ..., m\}$ *is a test set for* $L$. *The languages* $F_i$ *obviously satisfy properties* (c1) *and* (c2) *and therefore they are test sets for* $L$.

COROLLARY 12 (Test set construction). *For* $L$ *and arbitrary commutatively closed context-free language a test set* $F$ *for* $L$ *can be effectively found.*

This follows from conditions (c1) and (c2) for such an $F$. We also have

COROLLARY 13 (Test set decision). *For $L$ an arbitrary commutatively closed context-free language and $F$ an arbitrary commutatively closed finite set it is decidable whether or not $F$ is a test set for $L$.*

Finally, it should be mentioned that Theorem 9 can be generalized to include all $(1, 2)$ complete languages $L \subseteq \Sigma^*$ is $(1, 2)$ *complete* if for any pair of letters $a, b$ in $\Sigma$, $a \neq b$, such that $a$ occurs at least twice in a word of $L$ and there exists an $x$ in $L$ with $\{a, b\} \subseteq \text{alph}(x)$, then there is a $y$ in $\Sigma^*$ with *aaby*, *aayb*, *abay*, *abya*, *ayab* in $L$. This leads to an upper bound result on the size of the test sets, namely:

COROLLARY 14. *Let $L \subseteq \Sigma^*$ be $(1, 2)$ complete. Then there is a test set $F$ for $L$, such that $\#F \leqslant 2^n(n! + n) + 5n^2$, where $n = \#\Sigma$.*

*Proof.* It is easy to see that every $(1, 2)$ complete language $L \subseteq \Sigma^*$ contains a finite subset $F$ with the following properties:

(c1) For each $z$ in $L$ there exist $y_1, \ldots, y_m$ in $F$ and rational numbers $\alpha_1, \ldots, \alpha_m$ such that $\text{alph}(y_i) = \text{alph}(z)$ for $i = 1, \ldots, m$ and $\mathbf{p}(z) = \sum_{i=1}^{m} \alpha_i \times \mathbf{p}(y_i)$.

(d1) Let $z = x_0 a_1 x_1 a_2 x_2, \ldots, x_{t-1} a_t x_t$ be in $L$, $x_i$ in $\Sigma^*$, $a_i$ in $\Sigma$, such that the letters $a_i$ occur at most once in all words of $L$ and each letter in the $x_i$'s occurs at least twice in some word of $L$. Then for each such $z$ $F$ contains some word $y_0 a_1 y_1 a_2 y_2, \ldots, y_{t-1} a_t y_t$.

(d2) For each pair of letters $a, b$ in $\Sigma$, $a \neq b$, such that $a$ occurs at least twice in a word of $L$ and $\{a, b\} \subseteq \text{alph}(x)$ for some $x$ in $L$, $F$ contains five words *aaby*, *aayb*, *abay*, *abya*, *ayab* for some $y$ in $\Sigma^*$.

To show that $F$ is a test set for $L$, we carry over the case analysis in the proof of Theorem 9. Case 2 obviously can be covered by properties (d1), and (d2). In Case 1, we show first that $h_1(a)$, $h_2(a)$, $h_1(by)$, $h_2(by)$ commute as before. Now we represent $h_1(b)$, $h_1(y)$, $h_2(b)$, $h_2(y)$ in terms of the primitive root $u$ of $h_1(a)$. Inserting these in $h_1(aayb) = h_2(aayb)$, we derive easily that $h_1(a)$, $h_1(b)$, and $h_2(b)$ commute.

Since there are less than $2^n$ nonempty subalphabets of $\Sigma$, we need less than $2^n \times n$ words in $F$ to satisfy (c1). For (d1) and (d2) we need less than $2^n \times n!$ and $5n^2$ words, respectively. Thus we can find an $F$ with $\#F < 2^n(n! + n) + 5n^2$ satisfying properties (c1), (d1), and (d2). ∎

## REFERENCES

1. J. ALBERT, K. CULIK, II, AND J. KARHUMÄKI, "Test Sets for Context-Free Languages and Algebraic Systems of Equations over a Free Monoid," Tech. Report No. 104, Inst. für Angew. Inform. und Form. Beschr. verf., Univ. Karlsruhe, West Germany, 1981.
2. R. V. BOOK AND F.-J. BRANDENBURG, Equality sets and complexity classes, *SIAM J. Comput.* **9** (1980), 729–743.

3. K. CULIK, II, Some decidability results about regular and pushdown translations, *Inform. Process. Lett.* **8** (1979), 5–8.
4. K. CULIK, II, Homomorphisms: decidability, equality, and test sets, *in* "Formal Language Theory: Perspectives and Open Problems" (R. V. Book, Ed.), pp. 167–194, Academic Press, New York, 1980.
5. K. CULIK, II AND N. D. DIAMOND, A homomorphic characterization of time and space complexity classes of languages, *Internat. J. Comput. Math.* **8A** (1980), 207–222.
6. K. CULIK, II AND J. KARHUMÄKI, Systems of equations over a free monoid and Ehrenfeucht Conjecture, *Discrete Math.* (1982), to appear.
7. K. CULIK, II AND A. SALOMAA, On the decidability of homomorphism equivalence for languages, *J. Comput. System Sci.* **17** (1978), 163–175.
8. K. CULIK, II AND A. SALOMAA, Test sets and checking words for homomorphism equivalence, *J. Comput. System Sci.* **20** (1980), 379–395.
9. A. EHRENFEUCHT AND G. ROZENBERG, Elementary homomorphisms and a solution to the DOL sequence equivalence problem, *Theoret. Comput. Sci.* **7** (1978), 169–183.
10. S. GINSBURG, "Algebraic and Automata-Theoretic Properties of Formal Languages," North-Holland, Amsterdam, 1975.
11. M. A. HARRISON, "Introduction to Formal Language Theory," Addison–Wesley, Reading, Mass., 1978.
12. J. E. HOPCROFT AND J. D. ULLMAN, "Formal Languages and Their Relation to Automata," 2nd Ed., Addison–Wesley, Reading, Mass., 1980.
13. G. S. MAKANIN, The problem of solvability of equations in a free semigroup, *Mat. Sb.* **103** (145), (1977), 148–236. [Russian]
14. A. SALOMAA, "Formal Languages," Academic Press, New York, 1973.